新型冠状病毒（2019-nCoV）时空数据集及其典型应用

林浩，鲍君雅
（中国地质大学（武汉）地理与信息工程学院 湖北武汉 430074）

摘要：
[目的] 为广大科研工作者提供权威的、开放的和多尺度的新型冠状病毒（2019-nCoV）时空数据集。
[方法] 基于国家、各省市、地区卫健委的疫情通告进行数据采集、清洗、整理、制作和更新。
[结果] 2019-nCoV 时空数据集已完成数据共享，并保持 5 天一次的更新频率，可通过地址：https://github.com/Estelle0217/COVID-19-Epidemic-Dataset 获取。
[局限] 数据属性统计受限于官方疫情通告公布的信息，部分地区的地市级尺度缺少相应的统计信息。
[结论] 该数据集可以为疫情监测、防控、预测和预警提供重要的数据来源。

关键词：2019-nCoV；新型冠状病毒；数据集；多尺度；多时相
中图分类号：K901.2 **文献标识码：**A

1. 引言

自2019年12月以来，湖北省武汉市出现首例不明原因的肺炎患者（被诊断为感染新型冠状病毒，2019-nCoV），截止2020年2月5日24时，全球已有25个国家确诊感染2019-nCoV病毒累计28060例，死亡564例。2019-nCoV疫情目前正受到全球的生物学家[1-3]、医学家[4, 5]、流行病学家[6]和数理学家[7]等的广泛关注和研究。

病毒的传播通常具有时间和空间特性[8-10]。由于病毒具有一定的潜伏期[11]，即患者从感染病毒到发病往往需要一段时间；同时，潜伏期时患者的物理运动加速了病毒（或疫情）在地理空间上的传播和蔓延。记录疫情随时间和空间的变化信息（即时空信息）有助于理解疫情的传播速度和传播范围，从而为监测、预测和防控疫情提供重要的数据来源[12]。

当前，我国有关2019-nCoV疫情的数据主要由权威机构发布，包括国家卫生健康委员会（简称国家卫健委）和各省、自治区的卫生健康委员会（简称省卫健委）。然而，国家和省卫健委发布的数据或信息仅以文本通告，因该数据未与地理空间数据建立关联，不便于空间分析。虽然也有诸多平台（如丁香园[1]和人民网[2]等）通过爬取国家和省卫健委的数据实现2019-nCoV疫情的实时动态可视化，但是上述平台仅显示实时数据，较难获取疫情历史发展的时空数据。因此，本研

[1] https://ncov.dxy.cn/ncovh5/view/pneumonia
[2] http://society.people.com.cn/GB/369130/431577/index.html

究旨在为广大科研工作者提供权威的、开放的和多尺度的新型冠状病毒（2019-nCoV）时空数据集，有助于服务全球科学家为本次疫情建言献策。

本文的结构：第二章介绍2019-nCoV时空数据集的数据来源和特点；第三章介绍该时空数据集的潜在应用价值；第四章为结论与展望。

2. 数据介绍

新型冠状病毒（2019-nCoV）时空数据集的组织结构如图 1 所示。

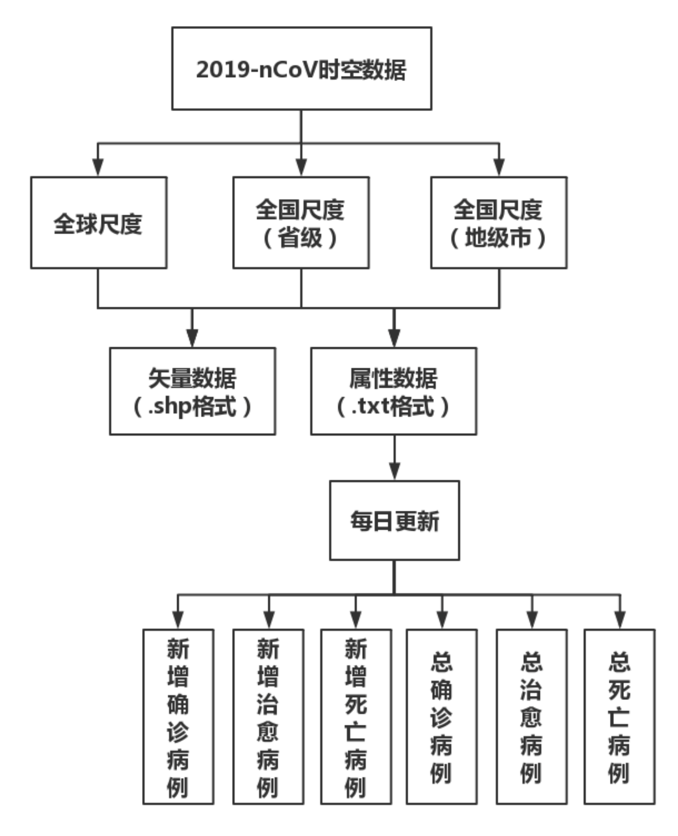


图 1 2019-nCoV 时空数据集的组织结构图

2019-nCoV 时空数据集[3]包括三种不同的尺度，即全球尺度、全国尺度（省级）和全国尺度（地级市）。各种尺度又包括矢量和属性数据。其中，矢量数据是指不同国家或地区的行政区划面状要素，主要包含全球国家、全国省级行政区划和全国地级市行政区划。其中，全国地级市行政区划由国家基础地理信息中心发布的全国基础地理数据库（现势性为 2015 年，比例尺为 1：100 万）中的县级行政区划面状要素经拼接、裁剪、重投影、融合所制得；全国省级行政区划沿用全国基础地理数据库中的省级行政区划面状要素；全球国家行政区划基于 GADM Ver. 3.6 （于 2018 年 5 月 6 日发布）的高分辨率面数据提取。矢量数据采用 WGS-84 大地坐标系，经纬度坐标，数据格式为 shapefile，可直接适用于 ERSI ArcGIS 等 GIS 平台；属性数据是指由权威机构发布的各个国家或地区的疫情数据（且每日更新），包括全球疫情数据、国内疫情数据（省级）和国内疫情数据（地市级），每份数据含 6 个统计表格：新增确诊病例、新增治愈病例、新

3 获取地址：https://github.com/Estelle0217/COVID-19-Epidemic-Dataset

增死亡病例、总确诊病例、总治愈病例和总死亡病例。每个统计表格包含字段：Region_CN、Region_EN 和 T+YYYY+MM+DD，其中在 T+YYYY+MM+DD 字段中，YYYMMDD 代表疫情数据的统计日期，字段属性代表当天的疫情数据（-1 则代表当天无数据）。数据格式为.txt 文本格式，可直接适用于 Microsoft Excel 和 ERSI Arcgis。

2.1 数据来源

- 矢量数据：包括全国和全球的矢量数据集。其中，全国矢量数据集又包括全国各个省的行政区划矢量数据和各个地级市的行政区划矢量数据，且它们均由国家基础地理信息中心提供，并可免费获取于全国地理信息资源目录服务系统平台（http://www.webmap.cn/commres.do?method=dataDownload）。全球矢量数据包括各个国家和地区的行政区划矢量数据。该数据获取于 Database of Global Administrative Areas(GADM:https://gadm.org/download_world.html)。该数据集是一个免费的、高精度的全球行政区划数据库[13]。
- 属性数据：包括全国和全球的属性数据。其中，全国属性数据是指全国各个省和地级市的疫情数据（包括新增确诊病例、新增治愈病例、新增死亡病例、总确诊病例、总治愈病例和总死亡病例六个字段）该数据获取于国家和省卫健委发布的 2019-nCoV 疫情通告。具体而言，省和地级市的病例数据获取于省卫健委的通告；香港、澳门和台湾的病例数据则获取于国家卫健委。全球属性数据是指各个国家的疫情数据，获取于人民网的疫情通告。属性数据的具体来源如表 1 所示。

表 1 2019-nCoV 时空数据集的属性数据来源

区域	数据来源	网址
省、直辖市和自治区	北京市卫健委	http://wjw.beijing.gov.cn/xwzx_20031/wnxw/
	天津市卫健委	http://wsjk.tj.gov.cn/col/col14/index.html
	河北省卫健委	http://wsjkw.hebei.gov.cn/index.do?templet=new_list&cid=14
	山西省卫健委	http://wjw.shanxi.gov.cn/wjywl02/index.hrh
	内蒙古自治区卫健委	http://wjw.nmg.gov.cn/xwzx/xwfb/index.shtml
	辽宁省卫健委	http://wsjk.ln.gov.cn/wst_wsjskx/
	吉林省卫健委	http://wjw.jlcity.gov.cn/gsgg/
	黑龙江省卫健委	http://wsjkw.hlj.gov.cn/index.php/Home/Zwgk/all/typeid/42
	上海市卫健委	http://wsjkw.sh.gov.cn/xwfb/index.html
	江苏省卫健委	http://wjw.jiangsu.gov.cn/col/col7290/index.html
	浙江省卫健委	http://www.zjwjw.gov.cn/col/col1202101/index.html
	安徽省卫健委	http://wjw.ah.gov.cn/news_list_477_1.html
	福建省卫健委	http://wjw.fujian.gov.cn/ztzl/gzbufk/yqtb/
	江西省卫健委	http://hc.jiangxi.gov.cn/xwzx/wjxw/index.shtml
	山东省卫健委	http://wsjkw.shandong.gov.cn/ztzl/rdzt/qlzhfkgz/tzgg/

	河南省卫健委	http://wsjkw.henan.gov.cn/channels/854.shtml
	湖北省卫健委	http://wjw.hubei.gov.cn/fbjd/dtyw/
	湖南省卫健委	http://wjw.hunan.gov.cn/wjw/xxgk/gzdt/zyxw_1/index.html
	广东省卫健委	http://wsjkw.gd.gov.cn/zwyw_yqxx/index.html
	广西壮族自治区卫健委	http://wsjkw.gxzf.gov.cn/zhuantiqu/ncov/ncovyqtb/
	海南省卫健委	http://wst.hainan.gov.cn/swjw/rdzt/yqfk/index.html
	重庆市卫健委	http://wsjkw.cq.gov.cn/yqxxyqtb/
	四川省卫健委	http://wsjkw.sc.gov.cn/scwsjkw/gzbd01/ztwzlmgl.shtml
	贵州省卫健委	http://www.gzhfpc.gov.cn/xwzx_500663/zwyw/
	云南省卫健委	http://ynswsjkw.yn.gov.cn/wjwWebsite/web/col?id=UU157976428326282067&cn=xxgzbd&pcn=ztlm&pid=UU145102906505319731
	西藏自治区卫健委	http://wjw.xizang.gov.cn/xwzx/wsjkdt/
	陕西省卫健委	http://sxwjw.shaanxi.gov.cn/col/col9/index.html
	甘肃省卫健委	http://wsjk.gansu.gov.cn/channel/10910/index.html
	青海省卫健委	https://wsjkw.qinghai.gov.cn/ztbd/yqjk/yqtb/index.html
	宁夏回族自治区卫健委	http://wsjkw.nx.gov.cn/yqfkdt/yqsd1.htm
	新疆维吾尔自治区卫健委	http://www.xjhfpc.gov.cn/ztzl/fkxxgzbdfygz/yqtb.htm
香港、澳门和台湾	国家卫健委	http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml
其它国家和地区	人民网	http://society.people.com.cn/GB/369130/431577/index.html

2.2 数据特点

本文将从权威性、完整性、时间精度和拓扑一致性四个方面阐述新型冠状病毒（2019-nCoV）时空数据集的特点。

2.2.1 权威性

全国的疫情数据获取于国家权威机构——国家和省卫健委；全球（其它国家）的疫情数据获取于人民网——国家权威互联网媒体平台。

2.2.2 完整性

矢量数据覆盖全球陆地范围，包括全球 253 个国家、地区及岛屿，中国 34 个省、直辖市和特别行政区，367 个地级市、州或地区。并按照全球（国家）、全国（省级行政区划）和全国（地级市行政区划）划分为三个不同的尺度。每个尺度的矢量数据记录了不同区域的中文（字段名：Region_CN）和英文（字段名：Region_EN）名称。

属性数据覆盖自 2019 年 12 月 31 日至今的各国、全国各省、市 2019-nCoV 疫情数据，包含新增确诊人数、新增治愈人数、新增死亡人数、总确诊人数、总治愈人数和总死亡人数共 18 个表格。由于 2019-nCoV 疫情正处于爆发阶段，国家卫生健康委员会发布的疫情数据存在一定的滞后性和误差，即总确诊人数、总治愈人数和总死亡人数存在核减或核增的情况，导致新增人数总和与总确诊人数不完全相等。此外，由于部分省份新增确诊人数存在地市级不明的情况，因此属性数据的统计均以卫健委最新公布的数据为准。

2.2.3 时间精度

该时空数据集的时间分辨率为 1 天，即统计国家和省卫健委当天发布的 0-24 时疫情数据。而且，该数据每天更新前一天发布的数据，直至国家和省卫健委不再发布有关 2019-nCoV 疫情的数据为止。需要注意的是：在疫情早期（即 2019 年 12 月 31 日至 2020 年 1 月 11 日期间），只有湖北省省份（本次疫情的主要来源地）的卫健委发布了疫情通告，因此上述时期内，其它省市卫健委的属性数据为空值，记录为-1。

2.2.4 拓扑一致性

为了保证全球和全国行政区划数据的边界一致性，本研究基于 ArcGIS 软件的“擦除工具”将全球行政区划的边界替换为全国基础地理数据库的国界线，避免了因拓扑错误而无法进行空间邻近分析。

3. 典型应用

3.1 属性数据分析

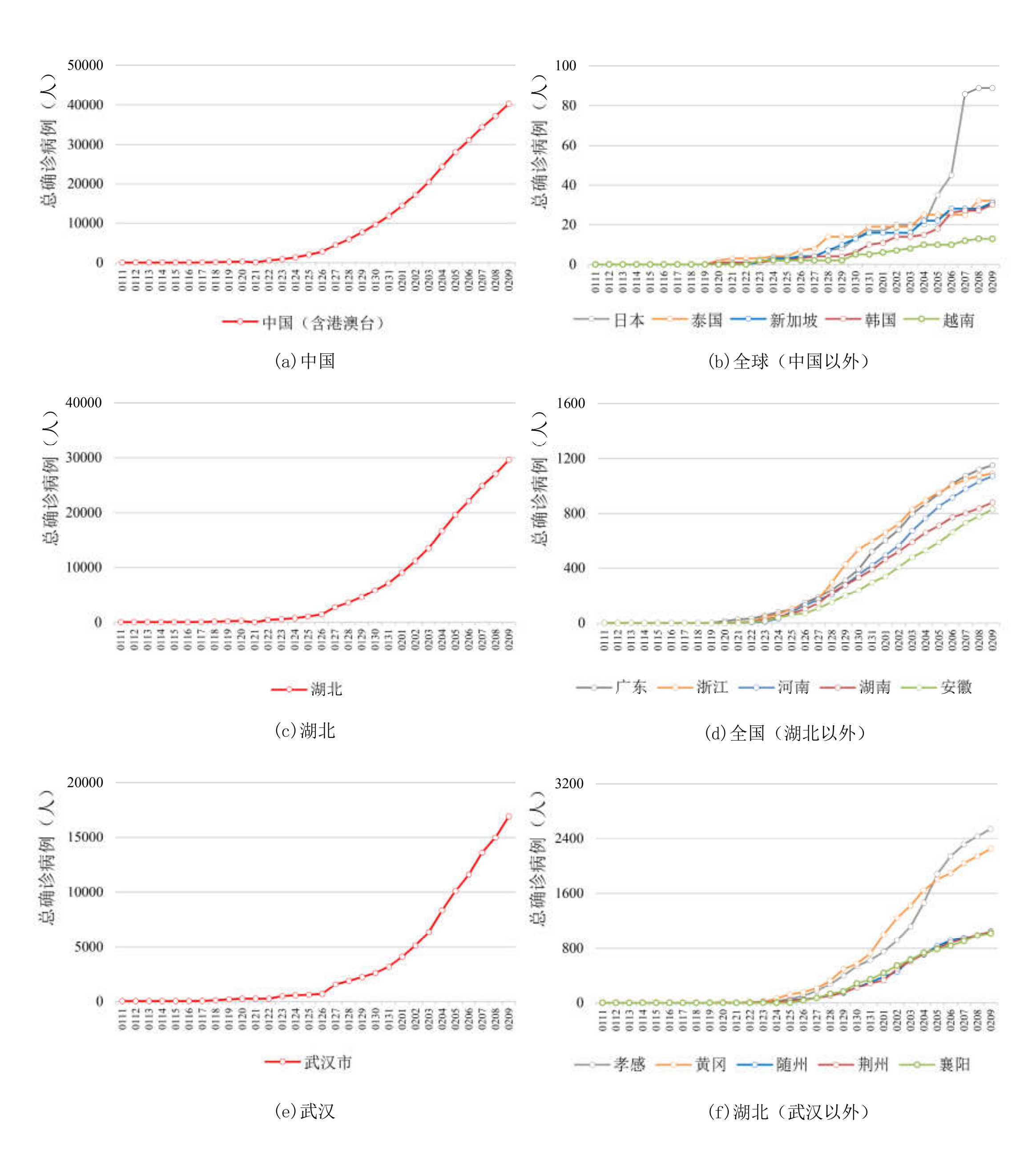


图 2 2019-nCoV 演化趋势

如图 2 所示为基于 2020 年 1 月 11 日至 2020 年 2 月 9 日时空数据集的属性数据所制得演化趋势，由图可以看出：

在国内方面（图 2a），总确诊病例于 1 月 26 日开始逐渐增多，截止至 2 月 9 日仍处于高值水平且还在持续增长中，反映出目前国内疫情仍处于爆发期阶段[14]。从全球（中国以外）范围看，国外总确诊病例数排行前五的国家（图 2b）分别是日本、泰国、新加坡、韩国和越南，除日本外其他国家的总确诊人数趋于平稳缓慢增长，其原因主要是：受“钻石公主”号邮轮（所属日本）集体感染 2019-nCoV 事件的影响，日本的总确诊人数增长较其他国家较快，而其他国家大都为散发的输入性病例[15]，少有集中感染病毒的案例，因此总确诊病例整体趋于平缓增长。

而在湖北省（图 2c）方面，总确诊病例远多于其他省份，尽管截至 2 月 9 日曲线出现有所缓和的迹象，但仍保持较快速率的增长。从全国（湖北省以外）

范围看，总确诊病例排行前五的省份（图 2d）分别是广东、浙江、河南、湖南和安徽，它们的总确诊病例增长趋势较为一致，增长的速率较疫情爆发初期逐渐趋于平缓，这意味着湖北省 13 个肺炎重灾区封城举措和省外的疾病防控工作取得了初步的成效。

在武汉（图 2e）方面，总确诊病例超过全省总确诊病例的 50%，并且仍以较快速率增长，反映出目前武汉市疫情形势不容乐观，防疫工作仍需要以武汉市作为重心。从湖北（地市级）范围看，总确诊病例排行前五的地级市（图 2f）分别是孝感、黄冈、随州、荆州和襄阳，其中，随州、荆州和襄阳的总确诊病例曲线自 2 月 2 日开始逐渐趋于稳定，孝感和黄冈总确诊病例虽仍在持续增加，但增加速率分别于 2 月 4 日和 2 月 6 日开始有所减缓。

3.2 时空数据分析

基于 ArcGIS 软件的“连接”工具，可直接将数据集中属性数据与矢量数据进行结合，从而得到疫情分布的多尺度、多时相制图和可视化。本研究以 2020 年 1 月 22 日至 2020 年 2 月 5 日的属性数据作为范例，制得全球（图 3）、全国省级（图 4）和全国地市级（图 5）三个不同尺度的疫情分布。由图可看出：

在全球范围内（图 3），2019-nCoV 的传播模式整体呈现以中国为辐射源，逐渐向周边国家及发达国家辐射的态势，且传播速度较为迅猛。在疫情初期（图 3a）仅有少数国家出现 2019-nCoV 确诊病例，如：韩国、日本、泰国和美国。随着疫情的传播，更多的国家如新加坡、俄罗斯、加拿大，澳大利亚、德国等出现确诊病例。其主要原因是：一方面，距离越近，地物间相关性越大[16]，因此中国周边国家受疫情影响严重；另一方面，虽然如澳大利亚、美国、欧洲各国等发达国家距离中国较远，但由于与中国的联系较为密切，因此疫情也同样较为严重。

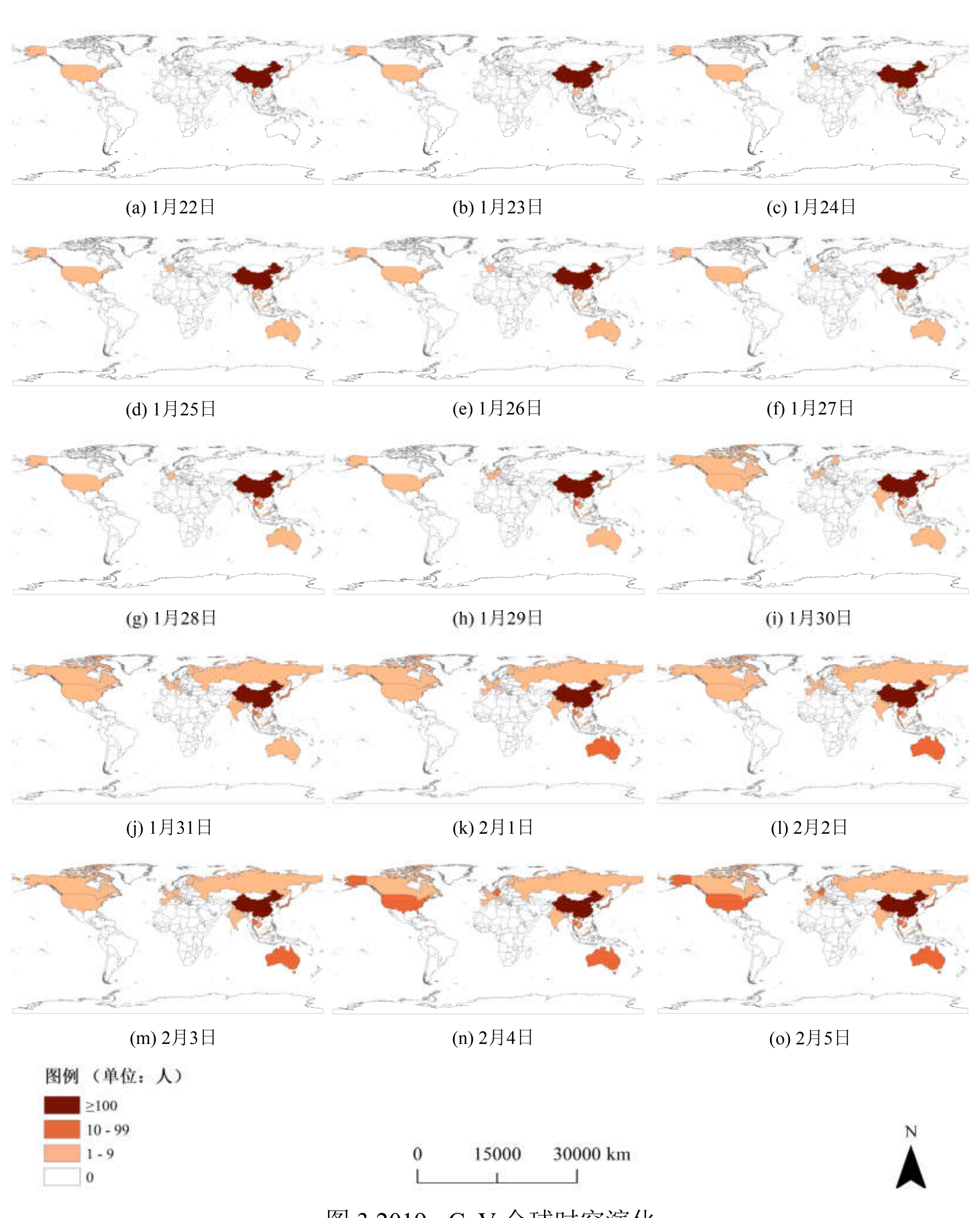


图 3 2019-nCoV 全球时空演化

在全国（省级）范围内（图 4），疫情整体主要发生在胡焕庸线以东的省份，其中，以广东、浙江、湖南、安徽等为代表的省份最为严重。在初期（图 4a），疫情主要爆发于北京、上海、浙江和广东地区。随后疫情开始以湖北省为中心向邻近省份扩散，如：江西、湖南、重庆、安徽、河南等地区的确诊人数均有明显增加。而在中后期（图 4o），疫情较为严重的地区则主要分布在中国中部和东南地区。

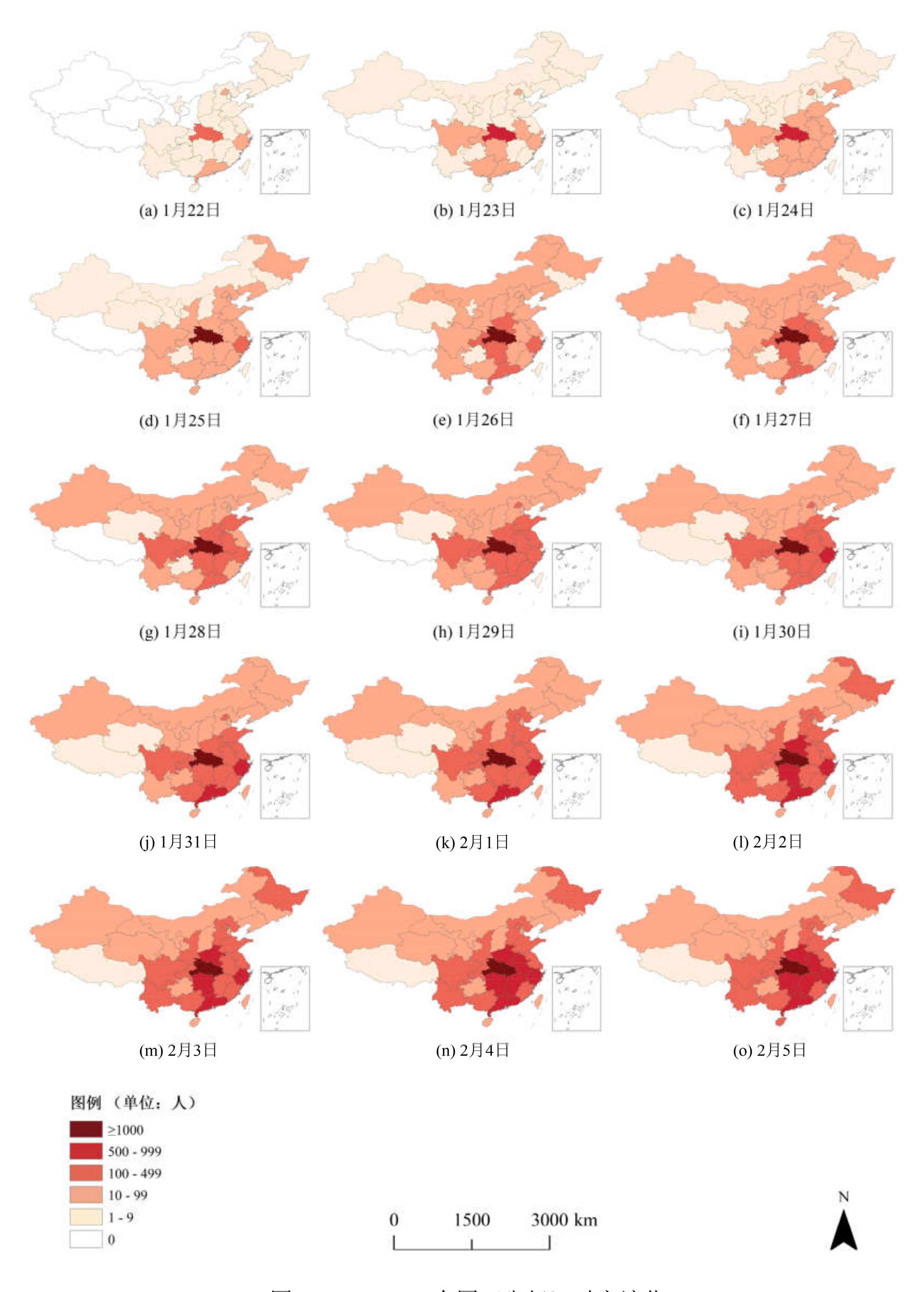


图 4 2019-nCoV 全国（省级）时空演化

在全国（地市级）范围内（图 5），考虑到武汉市于 1 月 23 日进行了封城措施，因此本研究将疫情时空演化主要分为武汉市、湖北省内和全国（湖北省以外）三个层次。在武汉市层面，由图可见疫情变化及其迅速，总确诊病例于 1 月 27 日已突破 1000 人。在湖北省内层面，疫情时空传播主要呈现以武汉作为辐射源向省内其他城市辐射的模式。在全国层面，北京、上海、广州和深圳为湖北省

外的疫情重灾区。出现上述现象的主要原因是：一方面，截止武汉市封城时已有超过 500 万人离开武汉，且大多数流向湖北省县市[4]，因此湖北省内地级市较其他省份地级市出现更多的病患；另一方面，北上广深等一线城市人口基数较大，人口流动性更高，因此确诊人数较湖北省以外的地级市更多。

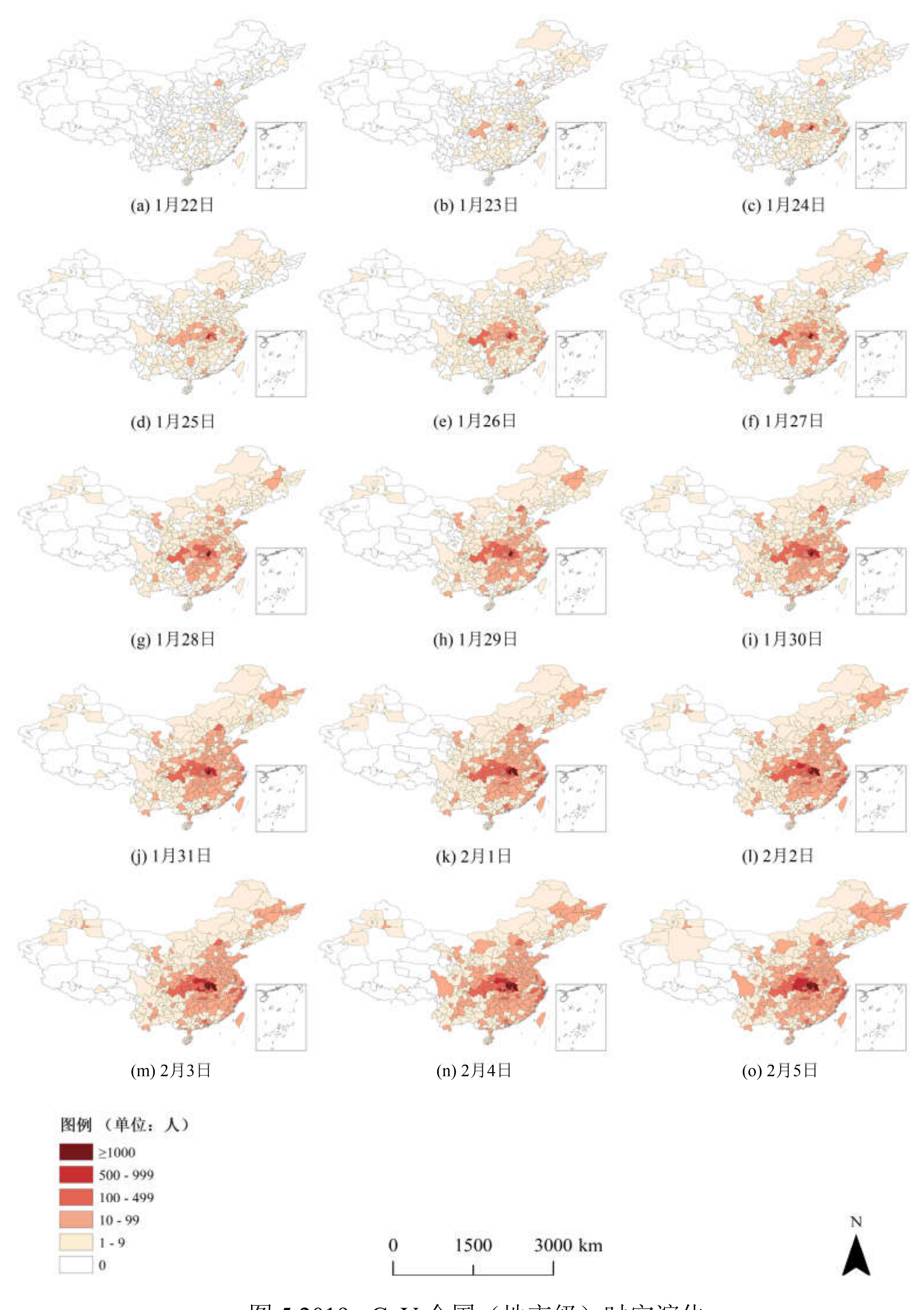


图 5 2019-nCoV 全国（地市级）时空演化

[4] http://qianxi.baidu.com/

4. 结论与展望

本研究向广大科研工作者提供一种权威的、开放的和多尺度的 2019-nCoV 时空数据集，以解决当前没有一个官方的渠道对 2019nCoV 疫情数据进行实时开源的问题。该时空数据集包括全球尺度、全国尺度（省级）和全国尺度（地级市）三种不同的尺度，每个尺度的数据记录了不同时相的疫情信息，包括：新增确诊病例、新增治愈病例、新增死亡病例、总确诊病例、总治愈病例和总死亡病例。此外，研究基于数据源进行属性及时空分析，论证了该数据集不仅可以为疫情监测、防控、预测和预警提供重要的数据来源，还能应用于 2019-nCoV 疫情的多尺度、多时态制图和可视化，为疫情的空间分布、演化、趋势分析和模拟预测提供指导。

参考文献：

[1] Li X, Zai J, Wang X, et al. Potential of large 'first generation' human-to-human transmission of 2019-nCoV[J]. Journal of Medical Virology, 2020.
[2] Corman V M, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR[J]. Eurosurveillance, 2020, 25(3).
[3] Benvenuto D, Giovannetti M, Ciccozzi A, et al. The 2019-new coronavirus epidemic: evidence for virus evolution [J]. Journal of Medical Virology, 2020.
[4] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China[J]. The Lancet, 2020.
[5] Chan J F, Yuan S, Kok K, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster[J]. The Lancet, 2020.
[6] Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia[J]. New England Journal of Medicine, 2020.
[7] Zhao S, Musa S S, Lin Q, et al. Estimating the Unreported Number of Novel Coronavirus (2019-nCoV) Cases in China in the First Half of January 2020: A Data-Driven Modelling Analysis of the Early Outbreak[J]. Journal of Clinical Medicine, 2020,9(2):388.
[8] Carroll M W, Matthews D A, Hiscox J A, et al. Temporal and spatial analysis of the 2014 - 2015 Ebola virus outbreak in West Africa[J]. Nature, 2015,524(7563):97-101.
[9] 钟少波. GIS和遥感应用于传染病流行病学研究[D]. 中国科学院研究生院（遥感应用研究所）, 2006.
[10] Zulu L C, Kalipeni E, Johannes E. Analyzing spatial clustering and the spatiotemporal nature and trends of HIV/AIDS prevalence using GIS: the case of Malawi, 1994-2010[J]. BMC infectious diseases, 2014,14(1):285.
[11] Rothe C, Schunk M, Sothmann P, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany[J]. New England Journal of Medicine, 2020.
[12] 郭凤云，路紫. 基于空间分析方法的疾病地理研究进展[J]. 地理信息世界, 2009,7(06):22-26.
[13] Hijmans R, Garcia N, Wieczorek J. Global Administrative Areas Database (GADM) Version 3.6[J]. UN: New York, NY, USA, 2018.
[14] Chen Z, Zhang W, Lu Y, et al. From SARS-CoV to Wuhan 2019-nCoV: Will History Repeat Itself? [J]. bioRxiv, 2020.
[15] Read J M, Bridgen J R, Cummings D A, et al. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions[J]. medRxiv, 2020.
[16] Tobler W R. A computer movie simulating urban growth in the Detroit region[J]. Economic geography,

1970,46(sup1):234-240.

(通讯作者:林浩 E-mail:cug_bird@cug.edu.cn)

作者贡献声明：

林浩：提出研究思路，设计研究方案；
鲍君雅：采集、清洗、整理、制作和更新数据；
林浩：数据应用与论文撰写；